6K

УДК 004.827

Ю.В. Крак, Д.В. Шкільнюк, К.С. Кручинін, Г.М. Єфімов
Інститут кібернетики імені В.М. Глушкова НАН України
Україна, 03680, м. Київ, проспект Академіка Глушкова, 40

Застосування аналізу зв'язних областей до задачі розпізнавання елементів дактильно-жестової мови

J.V. Krak, D.V. Shkilnjuk, K.S. Kruchinin, G.N. Efimov
V.M. Glushkov Institute of Cybernetic of NAS of Ukraine
Ukraine, 03680, Kiev, Glushkova ave., 40

Application of Analysis of Connected Regions to the Problem of Recognition of Elements of the Finger-Sign Language

Ю.В. Крак, Д.В. Шкильнюк, К.С. Кручинин, Г.Н. Ефимов
Институт кибернетики имени В.М. Глушкова НАН Украины
Украина, 03680, г. Киев, проспект Академика Глушкова, 40

Применение анализа связных областей к задаче распознавания элементов дактильно-жестового языка

Об'єктом дослідження є дактильно-жестова мова, яка використовується для спілкування людьми з вадами слуху. Метою дослідження є розробка та реалізація алгоритмів розпізнавання дактильно-жестової мови. У роботі розглядається застосування методів аналізу зв'язних областей, а також використання кутових параметрів для ідентифікації елементів дактильно-жестової мови.
Ключові слова: розпізнавання жестів, ідентифікація жесту.

The object of study is the finger-sign language used for communication by people with hearing impairments. The study is devoted to the development and implementation of algorithms for recognizing finger-sign language. This paper considers the application of methods for analysis of connected regions, and the use of angular parameters for identifying elements of finger-sign language.
Key words: gesture recognition, identification of gesture.

Объектом исследования является дактильно-жестовый язык, используемый для общения людьми с проблемами слуха. Целью исследования является разработка и реализация алгоритмов распознавания дактильно-жестового языка. В работе рассматривается применение методов анализа связных областей, а также использование угловых параметров для идентификации элементов дактильно-жестового языка.
Ключевые слова: распознавание жестов, идентификация жеста.

Вступ

Згідно з дослідженнями, близько 5% людей у світі мають ті або інші порушення слуху. На жаль, з кожним роком ця цифра зростає. У свою чергу порушення слуху породжує проблему розвитку мови спілкування цих людей. Більшість глухих є інвалідами з дитинства. Навчаючись у спецінтернатах, спілкуючись у власному середовищі, вони утворюють своєрідну субкультуру. Між собою такі люди спілкуються спеціальною жестовою мовою і їм досить важко контактувати з навколишнім світом.

Тому виникає необхідність створення комп'ютерних систем для вивчення жестової мови як інструменту комунікацій пересічних людей з людьми з обмеженим слухом.

Метою даної роботи є розробка загальної схеми роботи системи для розпізнавання елементів дактильної мови, а також використання спеціальних ознак положень кисті руки людини з метою підвищення ефективності розпізнавання жестів. Запропоновано ряд алгоритмів для ідентифікації окремих елементів жестів, в яких використовується аналіз кутових параметрів зображення і елементів дактиля. Даний підхід надає можливість правильно розпізнавати дактиль і уникнути певних труднощів при розпізнаванні станів кисті людини на різних фокусних відстанях.

Огляд існуючих підходів

Одним з найбільш ефективних підходів до розпізнавання елементів дактильної мови є використання механічних рукавиць. Так, в роботі [1] автори розробили систему розпізнавання американської дактильної абетки за допомогою використання механічної рукавички DataGlove. У роботі [2] описана система, яка заснована на нейромережному класифікаторі. Для вводу даних використовуються механічні рукавиці CyberGlove та Polhemus IsoTrack.

Широкого застосування при розпізнаванні елементів жестової мови отримали технології захвату рухів MoCap [3]. Для даної технології використовуються рукавиці, на яких розміщені перетворювачі інформації про положення від 8 до 32 камер.

Жест можна розпізнати за траєкторією руху руки на послідовності кадрів. У роботі [4] для розпізнавання жестів керування медіаплеєром або пакетом PowerPoint пропонується застосувати динамічну мережу Баєсса, яка включає приховану Марківську модель і фільтр Калмана. У роботі [5] розробляється система розпізнавання жестів, у якій використовуються сегментація зображення, відстеження обличчя і рук людини. Для вдалого відстеження обличчя і рук використовуються алгоритми виявлення шкіри, врахування особливостей руху, позицій рук і обличчя. Після виділення руки проводиться аналіз форми і порівняння з деякими еталонними зображеннями.

Схема роботи системи розпізнавання елементів дактильної мови

Ідентифікація елемента дактильної мови здійснюється за такою схемою (рис. 1). Розглянемо по кроках, як працює дана схема. Перед початком роботи системи завантажуються дані еталонних зображень, з якими у ході роботи порівнюється вхідне зображення. Також відбувається налаштування порогів освітлення та параметрів, що використовуються при побудові контурної моделі та скелета.

На вхід системі подається зображення через звичайну веб-камеру (рис. 2, а). Отримане зображення переводиться у формат HSV і за допомогою засобів сегментації відбувається пошук необхідних областей на зображенні відносно заданого порогу. Зображення, яке було отримане внаслідок сегментації, фільтрується і на виході маємо зображення з чітко виділеною рукою, на яку одягнена рукавичка (рис. 2, б).

Далі проводиться аналіз, до якого класу жестів відноситься вхідне зображення. Розпізнавання жестів можна визначити як відношення вхідних даних до певного класу образів за допомогою виділення суттєвих ознак або властивостей, які виділяють вхідні дані від загальної маси неістотних деталей. Під класом образів розуміється певна категорія, яка визначається рядом ознак, спільних для всіх її елементів [6].

У роботах [7], [8] було проведено класифікацію елементів дактильної мови. У результаті визначення, до якого класу відноситься вхідний елемент, подальші порівняння

з еталонними зображеннями будуть здійснюватися лише з елементами даного класу, а не з усіма елементами дактильної мови.

Система ідентифікації елементів дактильної мови працює наступним чином. Перед початком роботи формується таблиця з елементів дактильної мови і у відповідність кожному з них ставиться нуль. Відбувається розпізнавання за декількома алгоритмами, у результаті яких визначається елемент дактильної мови, до якого вхідне зображення найбільш схоже. Відповідно до отриманого елемента дактильної мови збільшується відповідний елемент таблиці на одиницю. Після закінчення роботи алгоритмів розпізнавання визначається елемент дактильної мови шляхом пошуку найбільшого елемента в таблиці, що була побудована на початку роботи.

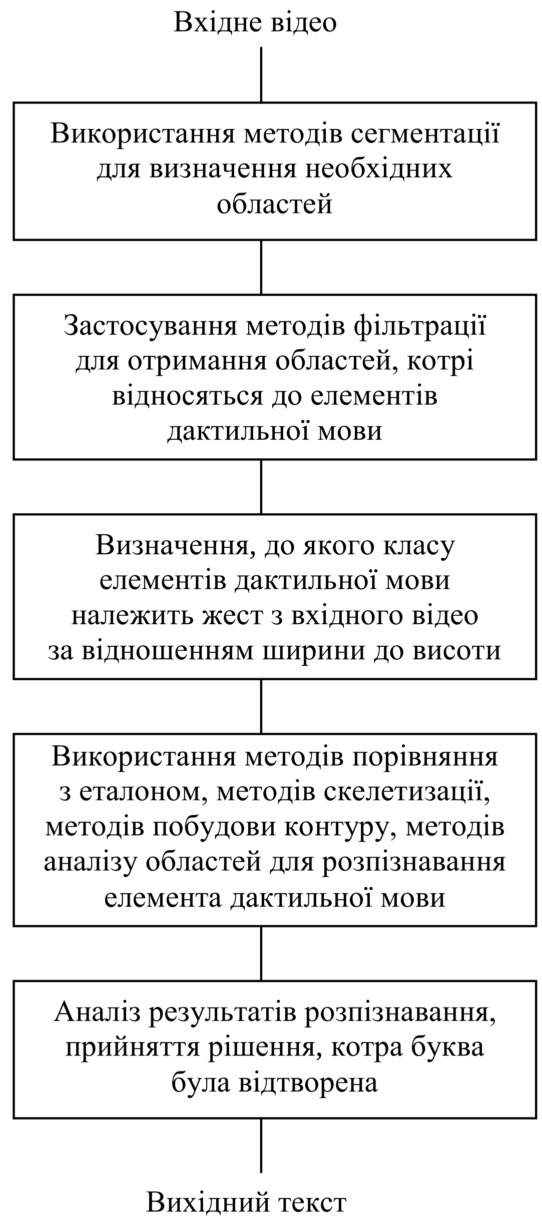


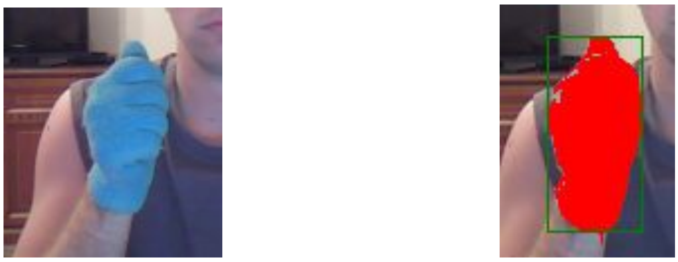
Рисунок 1 – Ідентифікація елемента дактильної мови

У випадку, коли в поле зору камери потрапляє жест з великими амплітудними рухами, необхідно відслідковувати траєкторію руху руки. Для відстеження траєкторії об'єкта необхідно слідкувати за координатами центра мас.

Розглянемо роботу даного методу на прикладі слова «ТРОЯ».

Слово «ТРОЯ» складається з чотирьох букв. Букви «Т», «Р», «О» відображаються нерухомими жестами, а буква «Я» відображається рухомим (рис. 3).

Зміни координат центра мас відобразимо на графіках. На вісі абсцис відображається момент часу, в який відображається жест. На вісі ординат відображаються координати x і y відповідно на першому і на другому графіках (рис. 4).



а) б)

Рисунок 2 – Дані, що отримані з веб-камери:
а) відтворення букви «А»; б) сегментація зображення



Рисунок 3 – Слово «Троя»

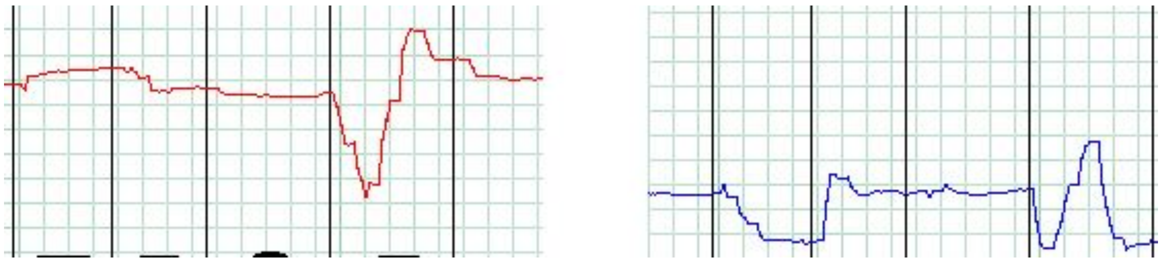


Рисунок 4 – Графічне подання зміни координат центра мас

Під час розпізнавання жестів відбувається побудова еталону вхідного жесту і порівняння з еталонами жестів (з бази знань) даного класу. Далі необхідно обчислити суму різниць між значеннями клітин еталонів і значенням клітин жесту, що розпізнається: $\sum_{i,j}^{n}\left|a_{ij} - M_{ij}^{k}\right| = A^k$, де a_{ij} – значення кожної клітини жесту, який розпізнається, M_{ij}^{k} – значення клітини кожного еталону. З множини A^k вибирається найменше. Найменше A^k визначить еталон, до якого відноситься жест, що розпізнається.

Зазначимо, що даний підхід застосовується і при розділенні зображення на вертикальні і горизонтальні смуги. У наслідок чого відбувається підрахунок різниць значень смуг еталонів зі значенням смуг жесту.

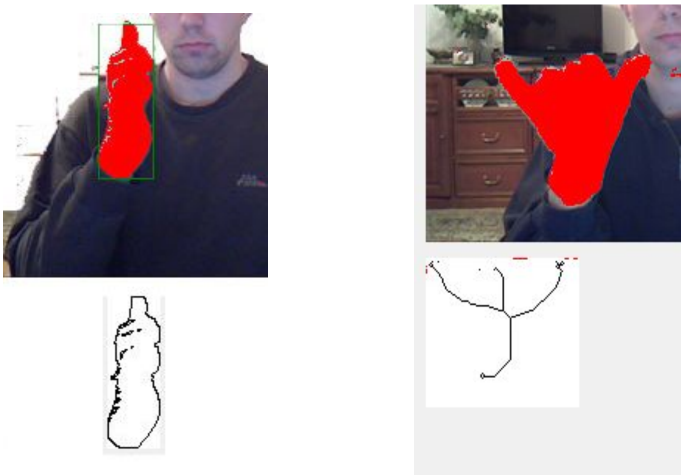
Після аналізу результатів наведеного алгоритму виникла необхідність у аналізі додаткових ознак жесту, а саме – аналізі контуру зображення [8] і скелета зображення.

Для побудови контуру були використані оператор Собеля і аналіз функції яскравості [9] вхідного зображення (рис. 5, а), а для побудови скелета зображення (рис. 5, б) було використано алгоритми Зонга-Сюня [10] та алгоритм «вогонь в прерії» [11].

Для більш ефективної ідентифікації елементів дактильної мови пропонується розглянути використання кутових параметрів кисті руки людини. Нехай N – кількість

пікселів, які відносяться до об’єкта. Всю множину пікселів $p(x, y)$, що відносяться до об’єкта, позначимо Ω. Тоді координати центра мас обчислюються наступним чином [7]:

$$x_c = \frac{1}{N} \sum_{p(x,y)\in\Omega} x,\; y_c = \frac{1}{N} \sum_{p(x,y)\in\Omega} y.$$



а) б)

Рисунок 5 – а) Побудова контуру жесту, що відтворює букву «Б»;
б) побудова скелета жесту, що відтворює букву «У»

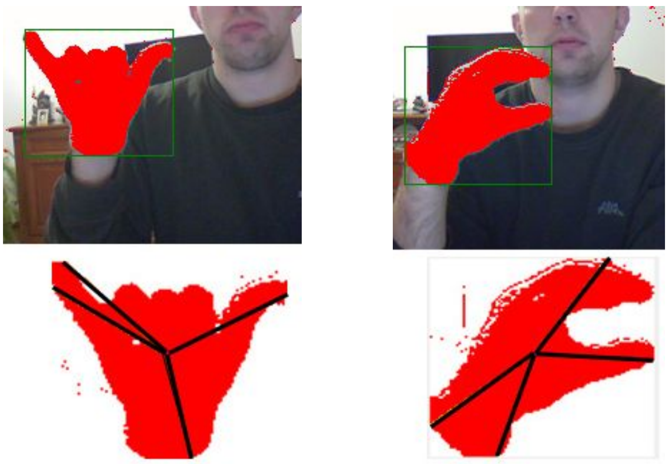
Проведемо з центра мас вектори у крайні точки області пікселів, які відносяться до кисті руки (рис. 6). Обрахуємо кути між побудованими векторами. Значення даних кутів будемо використовувати за ознаки для розпізнавання.

Розглянемо ще один спосіб виділення ознак для розпізнавання елементів дактильної мови. А саме знайдемо відношення довжин векторів, проведених з центра мас в крайні точки.

Нехай a_1, a_2, a_3, a_4 – довжини векторів, проведених з центра мас в крайні точки. Проведемо нормалізацію довжин векторів.

$$nA_1 = \frac{a_1}{a_1 + a_2 + a_3 + a_4},\; nA_2 = \frac{a_2}{a_1 + a_2 + a_3 + a_4},$$

$$nA_3 = \frac{a_3}{a_1 + a_2 + a_3 + a_4},\; nA_4 = \frac{a_4}{a_1 + a_2 + a_3 + a_4}.$$



Літера «У» Літера «С»

Рисунок 6 – Ідентифікація жесту за допомогою кутів між векторами,
проведених з центра мас в крайні точки

Отримані значення порівнюються з еталонними значеннями. На основі результатів порівняння вхідний жест класифікується до відповідного класу об'єктів.

Проведемо аналіз кутових параметрів для кутів, що утворені векторами, проведеними з центра мас в точки, які є кінцями ребер опуклої оболонки. Побудуємо опуклу оболонку навколо множини пікселів, які відносяться до об'єкта. Для цього необхідно побудувати контур області (позначимо його K) пікселів об'єкта. Наступний крок – це перевірка, чи дві довільні точки контуру утворюють ребро опуклої області. Цю перевірку можна зробити за правилом: для того щоб дві точки $(x_{c1}, y_{c1}) \in \mathrm{K}$ і $(x_{c2}, y_{c2}) \in \mathrm{K}$ утворювали ребро контуру, необхідно і досить, щоб

$$\forall (x_j, y_j) \in \mathrm{K},\ A*(x_j - x_{c1}) + B*(y_j - y_{c1}) \le 0,$$

або

$$\forall (x_j, y_j) \in \mathrm{K},\ A*(x_j - x_{c1}) + B*(y_j - y_{c1}) \ge 0,$$

де $A = y_{c1} - y_{c2}$, $B = x_{c2} - x_{c1}$.

Дану перевірку необхідно повторити для кожної точки, в результаті чого утвориться опукла оболонка.

Проведемо з центра мас вектори в кінці відрізків, які утворюють опуклу область (рис. 7). Обрахуємо кути між побудованими векторами. Значення даних кутів використовуються як ознака для розпізнавання.

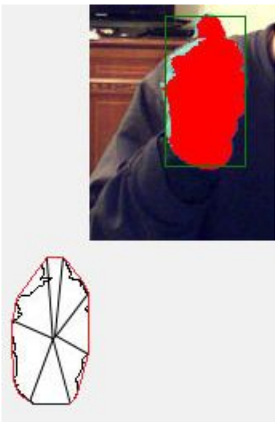


Рисунок 7 – Побудова опуклої оболонки і векторів з центра мас у відповідні точки

Проаналізувавши форму елементів дактильної мови та враховуючи те, що у всіх людей різні розміри руки, пропонується обрахувати наступні характеристики [12]: компактність; орієнтацію головної осі; видовження.

Компактність – це відношення квадрата периметра до площі області: $C = \frac{P^2}{S}$.

При роботі з зображенням площа – це кількість пікселів, які відносяться до області, периметр – кількість пікселів, котрі відносяться до контуру області. У роботі [8] описано ряд алгоритмів для побудови контурів елементів дактильної мови.

Для знаходження орієнтації головної осі та видовження необхідно обчислити дискретний центральний момент. Він знаходиться наступним чином:

$$m_{ij} = \sum_{(x,y)\in Reg} (x - \bar{x})^i (y - \bar{y})^j,\ \bar{x} = \frac{1}{n} \sum_{(x,y)\in Reg} x;\quad \bar{y} = \frac{1}{n} \sum_{(x,y)\in Reg} y,$$

де n – кількість пікселів, які відносяться до області.

Орієнтація головної осі обчислюється за такою формулою:

$$elongation = \frac{m_{20} + m_{02} + \sqrt{(m_{20} - m_{02})^2 + 4m_{11}^2}}{m_{20} + m_{02} - \sqrt{(m_{20} - m_{02})^2 + 4m_{11}^2}}.$$

Для визначення орієнтації (рис. 8) кисті руки можна скористатися наступною ознакою:

$$\theta = \frac{1}{2}\tan^{-1}\left(\frac{2m_{11}}{m_{20} - m_{02}}\right)$$



Рисунок 8 – Визначення орієнтації руки

Також пропонується використовувати більш складні інваріантні ознаки для ідентифікації елемента дактильної мови:

$$M_1 = m_{20} + m_{02}$$

$$M_2 = (m_{20} - m_{02})^2 + 4m_{11}{}^2$$

$$M_3 = (m_{30} - 3m_{12})^2 + (3m_{21} - m_{03})^2$$

$$M_4 = (m_{30} + m_{12})^2 + (m_{21} + m_{03})^2$$

$$M_5 = (m_{30} - 3m_{12})(m_{30} + m_{12})((m_{30} + m_{12})^2 - 3(m_{21} + m_{03})^2) +$$
$$+(3m_{21} - m_{03})(m_{21} + m_{03})(3(m_{30} + m_{12})^2 - (m_{21} + m_{03})^2)$$

$$M_6 = (m_{20} + m_{02})((m_{30} + m_{12})^2 - 3(m_{21} + m_{03})^2) +$$
$$+4m_{11}(m_{30} + m_{12})(m_{03} + m_{21})$$

$$M_7 = (3m_{21} - m_{03})(m_{12} + m_{30})((m_{30} + m_{12})^2 - 3(m_{21} + m_{03})^2) -$$
$$-(m_{30} - 3m_{12})(m_{12} + m_{03})(3(m_{30} + m_{12})^2 - (m_{21} + m_{03})^2).$$

У систему завантажуються значення ознак еталонних зображень. У ході роботи системи дані ознаки обчислюються для вхідного зображення і порівнюються з еталонними.

Висновки

У даній статті проаналізовано загальну схему роботи системи для ідентифікації елементів дактильної мови. Розглянуто роботу алгоритмів для розпізнавання жестів. Детально описано роботу алгоритмів розпізнавання за допомогою аналізу кутових параметрів, а саме розглянуто випадки роботи з крайніми точками та з опуклою оболонкою. У роботі описано використання таких ознак, як компактність, орієнтація головної осі, видовження.

На основі проведених випробувань і отриманих результатів розробляються алгоритми розпізнавання дактильно-жестової мови за відсутності рукавичок або за відсутності допоміжних міток на руках.

Література

1. Rung-Huei Liang. A Real-time Continuous Alphabetic Sign Language to Speech Conversion VR System / Rung-Huei Liang, Ming Ouhyoung // Computer Graphics Forum. – 1995. – Vol. 14, № 3. – P. C67-C76. (Also in EUROGRAPHICS’95, Holland).
2. Vamplew S. Recognition of Sign Language Using Neural Networks / S. Vamplew // Neurophyhological trends. – 2007. – Vol. 1. – P. 31-41.

3. Kitagawa M. MoCap for Artists. Workflow and Techiques for Motion Capture / M. Kitagawa, B. Windsor. – San Francisco : Focal Press is an imprint of Elsevier, 2008. – 216 p.
4. Suk H.-I. Robust Modelling and Recognition of Hand Gestures with Dynamic Bayesian Network / H.-I. Suk, B.-K. Sin, S.-W. Lee // Pattern Recognition, 2008. ICPR 2008. 19th International Conference, (8 – 11 Dec. 2008). – P. 1-4.
5. Real time hand gesture recognition including hand segmentation and tracking / T. Coogan, G.M. Awad, J. Han, A. Sutherland // In: ISVC 2006 : 2nd International Symposium on Visual Computing, (6 – 8 November 2006, Lake Tahoe, NV, USA).
6. Ту Дж. Принципы распознавания образов / Дж. Ту, Р. Гонсалес. – М. : Мир, 1978. – 411 с.
7. Крак Ю.В. Технологія розпізнавання елементів дактильно-жестової мови / Ю.В. Крак, Д.В. Шкільнюк // Штучний інтелект. – 2009. – № 3. – С. 564-572.
8. Крак Ю.В. Аналіз елементів дактильної жестової мови / Ю.В. Крак, Д.В. Шкільнюк // Штучний інтелект. – 2010. – № 3. – С. 322-328.
9. Форсайт Д. Компьютерное зрение. Современный подход / Д. Форсайт, Ж. Понс. – М. : «Вильямс», 2004. – 926 с.
10. Zhang T.Y. A fast parallel algorithm for thinning digital patterns / T.Y. Zhang, C.Y. Suen // Commun. ACM. – 1984. – V. 27, № 3. – P. 236-239.
11. Кауфе Ф. Взаимодействие робота с внешней средой / Кауфе Ф. – М. : Мир, 1985. – 279 с.
12. Режим доступу : http://courses.graphicon.ru/.

Literatura

1. Rung-Huei A. Computer Graphics Forum. P.C67-C76. Vol. 14. №. 3ю 1995. (also in EUROGRAPHICS'95. Holland).
2. Vamplew S. Neurophyhological trends. 2007. Mol. 1.P. 31-41.
3. Kitagawa M. MoCap for Artists. Workflow and Techiques for Motion Capture. San Francisco: Focal Press is an imprint of Elsevier. 2008.216 p.
4. Suk H.-I., Sin B.-K., Lee S.-W. Robust Modelling and Recognition of Hand Gestures with Dynamic Bayesian Network. Pattern Recognition. 2008. ICPR 2008. 19th International Conference. (8-11 Dec. 2008).P. 1-4.
5. Coogan T., Awad G. M., Han J., Sutherland A. Real time hand gesture recognition including hand segmentation and tracking. In: ISVC 2006 - 2nd International Symposium on Visual Computing. 6-8 November 2006. Lake Tahoe. NV. USA.
6. Tu J. Principy racpoznavanija obrazov. M.: Mir. 1978. 411 s.
7. Krak J.V. Technologija rospiznavannja daktylno-jestovoji movy. Shtuchnyj intelect. 2009. № 3.S. 564-572.
8. Krak J.V. Analiz elementiv daktylno-jestovoji movy. Shtuchnuj intelect. 2010. № 3. S. 322-328.
9. Forsait D. Kompjuternoje zrenije. Sovremennuj podchod.M.: Viljams. 2004. 926 s.
10. Zhang T. Y. Suen C. Y. Commun. ACM. 1984. V.27. № 3. P. 236-239.
11. Kaufe F.Vzaimodejstvie robota s vneshnej seredoj. M.: Mir. 1985. 279 s.
12. http://courses.graphicon.ru

Y.V. Krak, D.V. Shkilnuik, K.S. Kruchinin, G.N. Efimov

Application of Analysis of Connected Regions to the Problem of Recognition of Elements of the Finger-Sign Language

The problem of building of the system for elements identification and recognition in dactylology of the Ukrainian language are considered in the article. The basic idea is to design a mechanism analysis of the results of several recognition algorithms. An experimental study, which the person's depicted gestures with different sizes of hands in, have been made. It is proposed to solve the given problem with the following constraints: 1) hand, which plays a gesture, dressed in glove, with the color dramatically different from the external environment, 2) gesture played on low speed.

Identification of elements of the finger-sign language works as follows. Before using, the table of elements of the finger-sign language is formed and in accordance to each of

these elements, zero is put. Recognition is made by several algorithms. As the result, an element of the finger-sign language, which is more resembled to input image, is deteimined. According to the obtained element of the finger-sign language, the corresponding element in the table is increased by one. After work completion of algorithms of recognition, an element of the finger-sign language is defined by searching the largest element in the table that, whichwas built at the beginning of the work.

For identification and recognition of elements of the finger-sign language, the following scheme is used:

1. A video image is received to the input system. This image is transferred into HSV format, and by means of segmentation, the search for specific areas of the image relatively to the present threshold has place. The image obtained as a result of segmentation is filtered and output image is a distinct hand in that glove wearing;
2. Gestures, which class belongs the input image to, are analyzed;
3. An input image for multiple algorithms is recognized;
4. The result based on obtained consequence of algorithmsare formed.

Notably, library standards, which are used for recognition was built on basis of images of gestures that played by people with different sizes of hands.

On the basis of test results, the algorithms for recognition of the finger-sign language without gloves as well as without auxiliary labels on hands are developed.

Стаття надійшла до редакції 21.11.2011.